A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE aps

Research Article

Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning

Michael C. Frank, Noah D. Goodman, and Joshua B. Tenenbaum

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

ABSTRACT—*Word learning is a "chicken and egg" problem. If a child could understand speakers' utterances, it would be easy to learn the meanings of individual words, and once a child knows what many words mean, it is easy to infer speakers' intended meanings. To the beginning learner, however, both individual word meanings and speakers' intentions are unknown. We describe a computational model of word learning that solves these two inference problems in parallel, rather than relying exclusively on either the inferred meanings of utterances or cross-situational word-meaning associations. We tested our model using annotated corpus data and found that it inferred pairings between words and object concepts with higher precision than comparison models. Moreover, as the result of making probabilistic inferences about speakers' intentions, our model explains a variety of behavioral phenomena described in the word-learning literature. These phenomena include mutual exclusivity, one-trial learning, cross-situational learning, the role of words in object individuation, and the use of inferred intentions to disambiguate reference.*

When children learn their first words, they face a challenging joint-inference problem: They are both trying to infer what meaning a speaker is attempting to communicate at the moment a sentence is uttered and trying to learn the more stable mappings between words and referents that constitute the lexicon of their language. With either of these pieces of information, their task becomes considerably easier. Knowing the meanings of some words, a child can often figure out what a speaker is talking about, and inferring the meaning of a speaker's utterance allows a child to work backward and learn basic-level object names with relative ease. However, for a learner without either of these pieces of information, word learning is a hard computational problem. Quine (1960) suggested an apt metaphor: A word learner is climbing the inside of a chimney, "supporting himself against each side by pressure against the others" (p. 93).

Many accounts of word learning focus primarily on one aspect of this problem. Social theories suggest that learners rely on a rich understanding of the goals and intentions of speakers and assume that—at least in the case of object nouns—once the child understands what is being talked about, the mappings between words and referents are relatively easy to learn (St. Augustine, 397/1963; Baldwin, 1993; Bloom, 2002; Tomasello, 2003). These theories must assume some mechanism for making mappings, but this mechanism is often taken to be deterministic, and its details are rarely specified. In contrast, cross-situational accounts of word learning take advantage of the fact that words often refer to the immediate environment of the speaker, which allows learners to build a lexicon based on consistent associations between words and their referents (Locke, 1690/1964; Siskind, 1996; Smith, 2000; Yu & Smith, 2007).

Computational models of word learning have primarily followed the second, cross-situational strategy. Models using connectionist (Plunkett, Sinha, Møller, & Strandsby, 1992), deductive (Siskind, 1996), competition-based (Regier, 2005), and probabilistic (Yu & Ballard, 2007) methods have had significant success in accounting for many phenomena in word learning. However, speakers often talk about objects that are not visible and about actions that are not in progress at the moment of speech (Gleitman, 1990), adding noise to the correlations between words and objects. Thus, cross-situational and associative theories often appeal to external social cues, such as eye gaze (Smith, 2000; Yu & Ballard, 2007), but these are used as markers of salience (the "warm glow" of attention), rather than

Address correspondence to Michael C. Frank, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave., Room 46-3037D, Cambridge, MA 02139, e-mail: mcfrank@mit.edu.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

as evidence about internal states of the speaker, as in social theories.

More generally, cross-situational models address only one part of the learner's task. Such models are able to learn words, but they do not use the words that speakers utter to infer the speakers' intended meanings. By focusing only on the long-term mappings between items in the lexicon and referents in the world, purely cross-situational models treat the complex and variable communicative intentions of speakers as noise to be averaged out via repeated observations or minimized via the use of attentional cues, rather than as an important aspect of communication to be used in the learning task.

In this article, we present a model that captures both aspects of the word-learning task: It simultaneously infers what speakers are attempting to communicate and learns a lexicon. We first present the structure of the model and show that it obtains competitive results in learning from corpus data. We then show how the probabilistic structure of the model allows it to predict experimental results such as mutual exclusivity, one-trial word learning, and rapid cross-situational learning, as well as how its explicit representation of intention allows it to predict results on object individuation and the use of intentional cues.

DESIGN OF THE MODEL

Our model (which we refer to as the *intentional model*) consists of a set of variables representing the word-learning task and a set of probabilistic dependencies linking these variables in accordance with our assumptions about the task (see Fig. 1). The variables represent the lexicon of the language being learned, the referential intentions of the speaker, the words uttered by the speaker, and the learner's physical context at the time of the utterance. We define the relationships among these variables via two assumptions. The first is that what speakers intend to say is a function of the physical world around them. The second is that the words speakers utter are a function of what the speakers intend to say and how those intentions can be translated into the language they are speaking. With these assumptions and an observed corpus of situations—utterances and their physical context—our model can work backward using Bayesian inference to find the most likely lexicon.

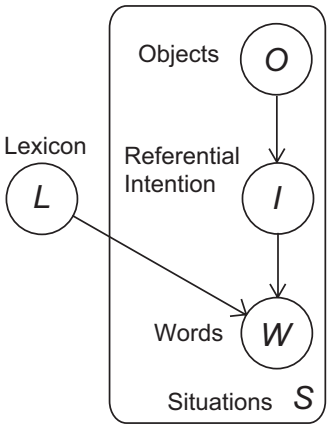

Fig. 1. Illustration of the dependence relations in our model. O, I, and W represent the objects present in the context, the objects that the speaker intends to refer to, and the words that the speaker utters, respectively. These variables are related within each situation s. The words that the speaker utters are additionally determined by the lexicon of the speaker's language, L, which does not change from situation to situation (and hence lies outside the representation of the set of situations).

Though a speaker's intentions could, in principle, be very complex, we limit ourselves here to the task of learning names for objects. Thus, we represent the physical context of an utterance as the set of objects present during the utterance, the speaker's referential intention as the object or objects he or she intends to refer to, and the lexicon as a set of mappings between words and objects. We also assume that objects are identified as instances of basic-level object categories, putting aside the challenge of identifying the particular aspect of an object being named (Xu & Tenenbaum, 2007).

Formally, our model defines a probability distribution over unobserved lexicons L and the observed corpus C of situations. Our goal is to infer the lexicon with the highest posterior probability. We find this posterior probability using Bayes' rule:

$$P(L|C) \propto P(C|L)P(L).$$

Bayes' rule factors the posterior probability of a lexicon given the corpus into two terms, the likelihood of the corpus given the lexicon and the prior probability distribution over lexicons. We chose a prior probability distribution that favored parsimony, making lexicons exponentially less probable as they included more word-object pairings: $P(L) \propto e^{-\alpha|L|}$. The choice of a simple prior puts most of the work of the model in the likelihood term, $P(C|L)$; hence, the likelihood term captures the learner's assumptions about the structure of the learning task.

The likelihood term can be written as a product over situations of the probability of the components of the corpus (the words W, objects O, and speaker's intentions I for each situation s), given the lexicon:

$$P(C|L) = \prod_{s \in C} P(W_s, O_s, I_s|L). \quad (1)$$

We can now use our assumptions about the structure of the task to factor Equation 1. First, W and O are conditionally independent given I (as shown in Fig. 1). Thus, we can rewrite the right-hand side as a product of $P(W_s|I_s,L)$, the probability of the words given the speaker's referential intentions and the lexicon, and $P(I_s|O_s)$, the probability of the speaker's intentions given the physical context. Second, because we cannot directly observe the speaker's referential intention, we sum over all possible values of I_s under the constraint that $I_s \subseteq O_s$ (i.e., that the relevant subset of possible intentions consists of those that refer to a subset of the objects in the physical context). Because speakers often refer to objects outside the physical context, I_s

can be empty (Gleitman, 1990). We rewrite Equation 1 as

$$P(C|L) = \prod_{s \in C} \sum_{I_s \subseteq O_s} P(W_s|I_s, L) \cdot P(I_s|O_s).$$

For simplicity (and because we have no information about I_s other than the words that are uttered), we set $P(I_s|O_s) \propto 1$ so that all possible intentions are equally likely.

To complete our definition of the model, we define the term $P(W_s|I_s,L)$ by assuming that the words $\{w_1 \ldots w_n\}$ in W_s are generated independently (ignoring any syntax), and that there are two possible causes for uttering a word. A word is uttered either *referentially*—in order to refer to an object in the speaker's intention set—or *nonreferentially*. The probability of a word being uttered if it is used referentially (P_R) is the probability that it will be chosen from the lexicon to refer to any of the intended referents. The probability of a word being uttered if it is used nonreferentially (P_{NR}) is the probability that it will be picked from the lexicon at random, independently of the speaker's referential intention. Verbs, adjectives, and function words are generated nonreferentially, as are object nouns for which the relevant object is not currently present. The parameter γ is the probability that a word is used referentially in any given context. Thus, we have

$$P(W_s|I_s, L) = \prod_{w \in W_s} \left[\gamma \cdot \sum_{o \in I_s} \frac{1}{|I_s|} P_R(w|o, L) + (1 - \gamma) \cdot P_{NR}(w|L) \right].$$

The probability of a word being used referentially for an object, $P_R(w|o,L)$, is the probability that the word is chosen uniformly from the set of words linked to that object in the lexicon. If there are, for example, two words linked to an object in the lexicon, each word has a probability of .5 of being used to refer to that object; if a word is not linked to an object, its referential probability for that object is 0. The nonreferential probability of a word being used is the probability of its being picked from the full set of words observed in the corpus. This choice is made with probability proportional to 1 if the word is not in the lexicon and with probability proportional to κ otherwise. Thus, when κ is less than 1, words that are in the lexicon are less likely to be uttered nonreferentially than words that are not in the lexicon.

In the simulations reported here, we employed stochastic search methods using simulated tempering (Marinari & Parisi, 1992) to find the lexicon with the maximum a posteriori probability given an observed corpus. In our on-line supporting information (see p. 585), we provide code for all models discussed here (Word-Learning Model Code) and implementational details regarding the search methods and simulations (Technical Appendix).

CORPUS EVALUATION

Method

We coded two video files (me06 and di03, each approximately 10 min long) from the Rollins section of the Child Language Data Exchange System (CHILDES; MacWhinney, 2000). In these videos, two preverbal infants and their mothers played with a set of toys. Each line of the transcripts was annotated with a list of all midsize objects judged to be visible to the infant.[1]

For comparison, we implemented several other models of cross-situational word learning using co-occurrence frequency, conditional probability, and point-wise mutual information. We also implemented IBM Machine Translation Model I (Brown, Pietra, Pietra, & Mercer, 1994), the statistical machine-translation model used by Yu and Ballard (2007). We used the translation model to compute association probabilities both for objects given words (as in Yu & Ballard) and for words given objects.

We evaluated all models both on the accuracy of the lexicons they learned and on their inferences regarding the speakers' intent. Each of the comparison models produced a single summary statistic linking words and objects (e.g., association probability). We chose the threshold value for this statistic that maximized the F score—the harmonic mean of precision (proportion of pairings that were correct) and recall (proportion of total correct pairings that were found)—of the resulting lexicon. We then used each model to make guesses about the speaker's intended referents for each utterance. For our model, we chose the intention with the highest posterior probability given the best lexicon; for the comparison models, we assumed that the intended referents were those objects for which the matching words in the best lexicon had been uttered. We computed scores relative to a gold-standard lexicon and a gold-standard set of intents, both created by a human coder. The gold-standard lexicon incorporated all standard word-object pairings (for a lamb toy, "lamb"), plurals ("lambs"), and baby talk ("lambie"); the gold-standard set of intents contained the human coder's best judgment of which objects in the visual context were being referred to in each utterance.

Results

Our intentional model substantially outperformed the comparison models with respect to the lexicons the model learned (Table 1) and the intentions it inferred (Table 2). This advantage was robust across systematic variation of the model's three free parameters (α, the degree to which the model favors small lexicons; γ, the probability of using words referentially; and κ, the parameter controlling how likely words in the lexicon were to be used nonreferentially, compared with words outside the lexicon). In addition, the advantage remained when κ and γ were set to their maximum a posteriori values (the joint empirical Bayes estimate; see Carlin & Louis, 1996), reducing the number of free parameters to one—the same number as in the baseline models. Table 3 shows the best lexicon learned by the intentional model.

[1]These videos are the same ones used by Yu and Ballard (2007); the annotations are our own.

TABLE 1

Precision, Recall, and F *Score of the Best Lexicon Found by Each Model When Run on the Annotated Data From the Child Language Data Exchange System*

Model	Precision	Recall	*F* score
Association frequency	.06	.26	.10
Conditional probability (object|word)	.07	.21	.10
Conditional probability (word|object)	.07	.32	.11
Mutual information	.06	.47	.11
Translation model (object|word)	.07	.32	.12
Translation model (word|object)	.15	.38	.22
Intentional model	**.67**	**.47**	**.55**
Intentional model (one parameter)	.57	.38	.46

Note. The highest values obtained are highlighted in boldface (differences between values may not be apparent because of rounding).

TABLE 2

Precision, Recall, and F *Score for the Referential Intentions Found by Each Model, Using the Lexicons Scored in Table 1*

Model	Precision	Recall	*F* score
Association frequency	.27	**.81**	.40
Conditional probability (object|word)	.59	.36	.45
Conditional probability (word|object)	.32	.79	.46
Mutual information	.36	.37	.37
Translation model (object|word)	.57	.41	.48
Translation model (word|object)	.40	.57	.47
Intentional model	**.83**	.45	**.58**
Intentional model (one parameter)	.77	.36	.50

Note. The highest values obtained are highlighted in boldface.

Both the simple statistical models and the translation model found a large number of spurious lexical items; the best lexicons found by these models were considerably larger than the best lexicon found by our model.[2] The high precision of the lexicon found by our model was likely due to two factors. First, the distinction between referential and nonreferential words allowed our model to exclude from the lexicon words that were used without a consistent referent. Second, the ability of the model to infer an empty intention allowed it to discount utterances that did not contain references to any object in the immediate context.

PREDICTION OF EXPERIMENTAL RESULTS

As a consequence of its structure, our model exhibits a graded preference for certain kinds of lexicons and utterance interpretations. First, the model prefers sparse lexicons because the simplicity prior biases the model against adding word-object mappings that do not increase the likelihood of the data. Second, the model tends to prefer one-to-one lexicons if they are consistent with the observed data, because having multiple words that can refer to an object reduces the probability of any single word being used consistently to refer to that object. Finally, the model prefers that people have intentions to talk about the objects that are present, because words that are generated referentially from an intention to talk about an object have higher likelihood than words that are generated nonreferentially at random from the entire vocabulary of the language. These three preferences allow the model to predict a number of empirical results in early word learning.

[2]The performance we report for the translation model is considerably lower than that reported by Yu and Ballard (2007). Several factors may have contributed to this difference. The speech transcripts used in our study were taken directly from CHILDES, whereas those in Yu and Ballard's study may have differed because the utterance boundaries in their transcripts were identified via the application of speech-recognition software to the original videos (C. Yu, personal communication, June 25, 2007). Also, our coding of the corpus and theirs may have included slightly different sets of objects for each situation. Finally, our gold-standard lexicon likely differed from theirs as well.

Cross-Situational Word Learning

Recent work has provided strong evidence that both adults and children are able to learn cross-situational associations between words and objects even in the absence of individually unambiguous trials (Smith & Yu, 2008; Vouloumanos, 2008; Yu & Smith, 2007). Our model and all of the comparison models successfully found all the correct word-object pairings with perfect precision and recall when presented with the stimulus

TABLE 3

The Best Lexicon Found by the Intentional Model

Word	Object
bear	**bear**
bigbird	**bird**
bird	**duck**
birdie	**duck**
book	**book**
bottle	bear
bunnies	**bunny**
bunnyrabbit	**bunny**
hand	**hand**
hat	**hat**
hiphop	mirror
kittycat	**kitty**
lamb	**lamb**
laugh	cow
meow	baby
mhmm	hand
mirror	**mirror**
moocow	**cow**
oink	pig
on	ring
pig	**pig**
put	ring
ring	**ring**
sheep	**sheep**

Note. Entries judged to be correct according to the gold standard are shown in boldface.

materials used in Yu and Smith's artificial-language experiments. Given that the statistics in these experiments sharply favor the correct lexicon, the success of human learners does not help to differentiate among the models we compared.

Mutual Exclusivity

In classic demonstrations of mutual exclusivity, a child is presented with two objects, one familiar and one novel. The experimenter asks, "Can you hand me the [novel name]?" and the child hands over the novel object, indicating that he or she has correctly inferred that the novel name refers to it (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Markman & Wachtel, 1988). Markman and her colleagues (Markman, 1989; Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003) have suggested that children possess a *principle of mutual exclusivity* that leads them to prefer lexicons with only one label for each object. Other researchers have suggested alternate explanations, including more limited principles that are learned with experience (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Mervis & Bertrand, 1994) and more general pragmatic principles (Clark, 1988, 2002).

Without building in an explicit assumption of mutual exclusivity, our model shows a soft preference for one-to-one mappings. We tested our model in the classic mutual-exclusivity paradigm (Markman & Wachtel, 1988) and found that it correctly inferred that the novel word mapped to the novel object. We scored four possible lexicons on our original CHILDES corpus extended with a mutual-exclusivity scenario (the word "dax" is uttered in the presence of a bird toy and a novel object, a dax). Each panel of Figure 2 shows the relative posterior probability of one lexicon under the model along with its relative prior probability and the relative likelihood it assigns to both the original corpus data and the new mutual-exclusivity situation.

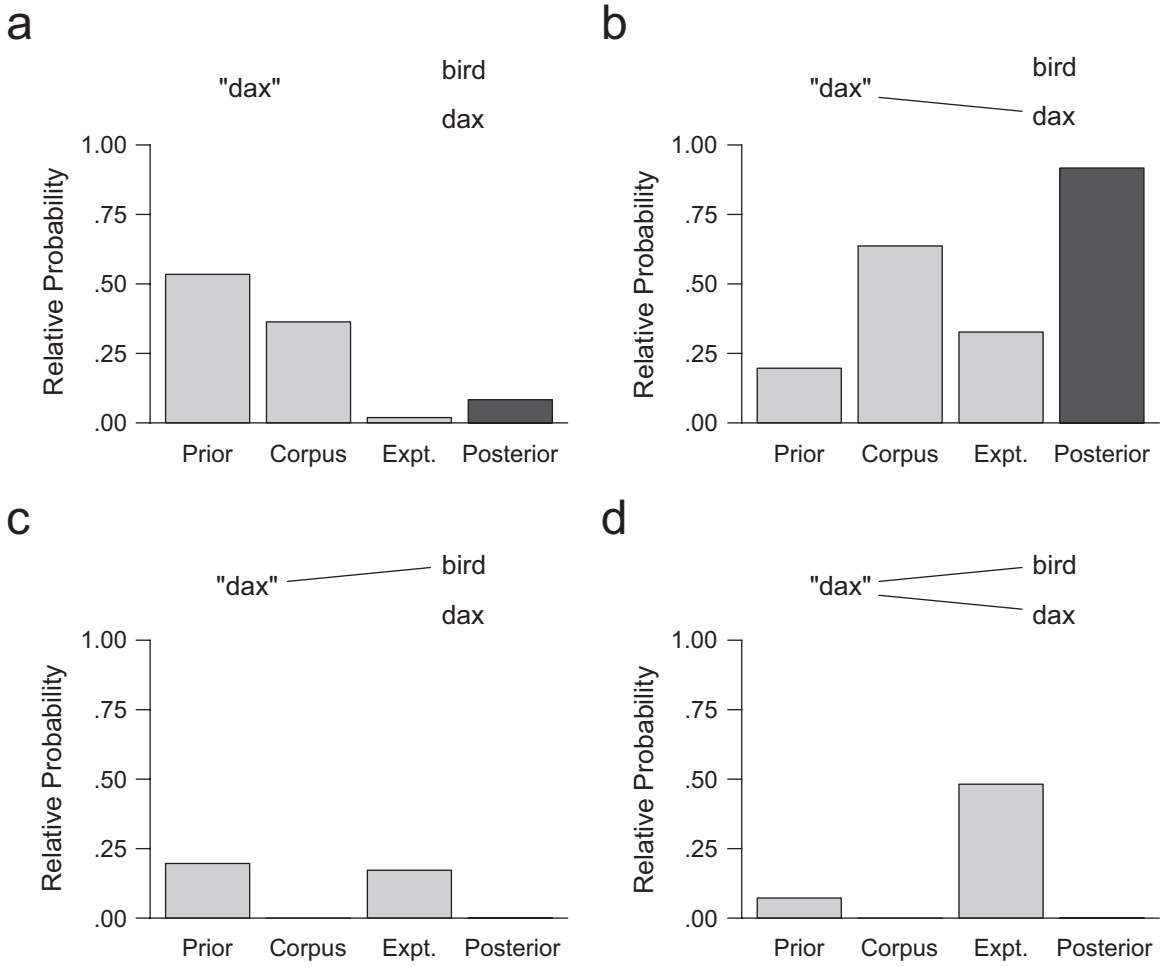

Fig. 2. Our model's relative probabilities for a learner's possible lexical hypotheses in a mutual-exclusivity experiment. If the experimenter utters the novel word ''dax'' in the presence of a novel object (a dax) and a known object (a bird), the learner can decide the word refers to (a) neither object, (b) the dax, (c) the bird, or (d) both objects. Each panel shows the prior probability, the likelihood of the original corpus data, the likelihood of the experimental situation, and the posterior (total) probability for one hypothesis. A parameter set less extreme than the one in the corpus simulations was used so that absolute rather than log probabilities could be plotted, but this change did not affect the ordering of hypotheses. All probabilities were normalized across the four possible hypotheses.

Lexicons in which "dax" was mapped to the familiar object, bird (Figs. 2c and 2d), were unlikely with respect to the original corpus because each sentence in which "bird" was uttered became less likely as a result of the unrealized possibility of hearing "dax." The lexicon in which no new words were learned (Fig. 2a) had a higher prior probability because it involved no growth in the size of the lexicon, but had a low likelihood in the experimental context (because under this hypothesis, the word "dax" was not in the lexicon and hence must have been uttered nonreferentially, rather than being uttered because of an intention to refer to the dax). Overall, our model preferred the correct lexicon (Fig. 2b).

This result is not unique to our model: The basic finding of mutual exclusivity is captured by many of the baseline models we tested. In the example just discussed, the conditional probability of the word "dax" given the presence of the bird is quite low, whereas the probability of the word "bird" given the presence of the bird is still very high. Combined with the demonstration that adults and infants are able to use some sort of statistical information in cross-situational learning tasks (Smith & Yu, 2008; Yu & Smith, 2007), the success of our model and others suggests that it is not necessary to posit domain-specific principles to account for findings of mutual exclusivity.

One-Trial Learning

Another classic result in the literature on word learning is the ability of children to learn a new word from only one or a small number of incidental exposures (Carey, 1978; Markson & Bloom, 1997). Our model and the comparison models predict that there are some situations that—in conjunction with the learner's previous experiences—can provide sufficient evidence for a word's referent to be inferred after a single exposure; in fact, the mutual-exclusivity experiment just described provides one such situation. We next turn to a set of experimental results that, to the best of our knowledge, cannot be captured by the comparison models.

Object Individuation

Even before their first birthday, infants are able to use the presence of words as an aid in individuating objects (Xu, 2002). In one experiment by Xu, infants saw first a duck and then a ball emerge from and then retreat behind a screen. Infants in the two-word condition heard, "look, a duck" and then "look, a ball"; infants in the one-word condition heard, "look, a toy" twice. At test, the screen dropped, revealing either one or two objects. Infants in the one-word condition looked longer when two objects were revealed than when one object was revealed (indicating that they expected only one object), whereas infants in the two-word condition looked slightly longer when one object was revealed than when two objects were revealed (indicating that they expected two objects and were surprised that one had disappeared).

Why would hearing two different labels allow infants to make the inference that two different objects were behind the screen? Perhaps the infants' assumptions about how words are normally used allowed them to infer what state of the world (one object or two) would be most likely to make a speaker utter the labels they heard. Because our model prefers lexicons with more one-to-one mappings and lexicons that interpret the corpus as having more referential words, the best interpretation of Xu's (2002) two-word condition in our model is that each word refers to a different object and that both words are being used referentially. Under this interpretation, there must be two different objects behind the screen, so that each of the two words can be used referentially to refer to one of the objects. Likewise, in the one-word condition, the most likely interpretation is that the one word refers to one object and that the word is being used referentially for that object; thus, there is likely only one object behind the screen.

To simulate Xu's (2002) paradigm formally in our model, we created sets of situations corresponding to the two experimental conditions. For each set, we created two construals: one in which there were two objects (though they were seen one at a time) and one in which there was only one object. To simulate the infant's uncertainty about the meanings of the word or words in the experiment, we evaluated each construal for all possible lexicons. In order to link the probability of a particular state of the world under our model to the looking time of infants in Xu's study, we used surprisal (negative log probability), a measure that has previously been used successfully to link model probabilities to human reaction time data (Levy, 2007). We compared the surprisal of the model for the two construals of each experimental condition (e.g., two words, one object vs. two words, two objects). This comparison can be interpreted as measuring, for a learner with no knowledge of what the words mean, how much more or less surprising it would be to find one object as opposed to two behind the screen. We found a crossover interaction, with surprisal being higher when the number of words did not match the number of objects. This result mirrors those Xu (2002) obtained (see Fig. 3). Thus, our model was able to use its assumptions about how words work to make inferences about the states of the world that caused a speaker to produce particular utterances.

Intention Reading

Baldwin (1993) conducted an experiment in which 19-month-old toddlers were shown two opaque containers, each containing a different novel toy. The experimenter opened one container, named the toy inside without showing the child the contents of the container, gave the child the toy from the second container to play with, and finally gave the child the first (labeled) object. Despite the greater temporal contiguity between the label and the second toy, the children showed evidence of learning that the label corresponded to the first toy. Baldwin interpreted these results as evidence that the children used the experimenter's

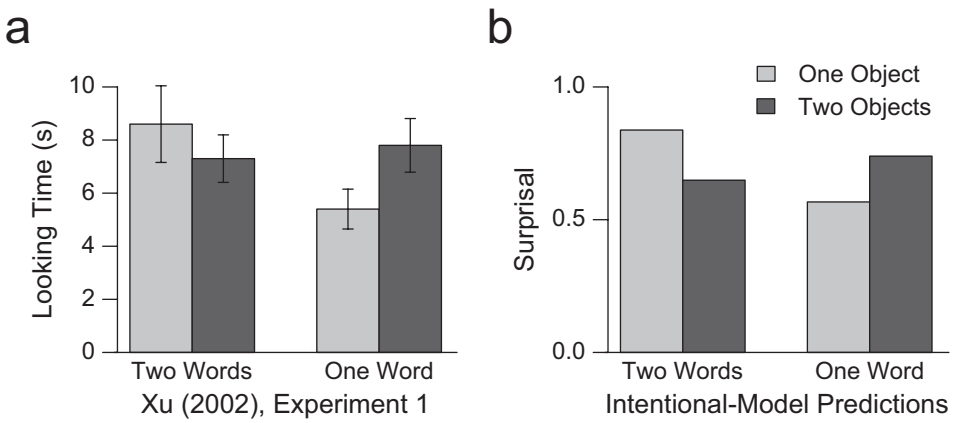

Fig. 3. Looking-time results reflecting infants' use of labels to individuate objects and simulation results from the intentional model proposed here. The graphs present (a) mean looking time in Xu's (2002) Experiment 1 as a function of experimental condition and (b) surprisal (negative log probability) calculated from the model in the same four conditions.

referential intention as their preferred guide to the meaning of the novel label. Our model, built around inferring the speaker's intended referents, can capture this interpretation directly. To illustrate this point, we constructed a situation with two novel objects and a single novel word. Whereas we previously treated the speaker's intention as a hidden variable, to model Baldwin's task, we gave the model additional information that the speaker intended to refer to the first novel object. The model then highly preferred the correct pairing.

This result should not be surprising, as we directly incorporated the referential intention of the speaker into our simulation. But a model that does not incorporate a representation of referential intent will be unable to predict Baldwin's (1993) results. Under a salience view, in order for the correct mapping to occur, the object that is out of sight would have to be more perceptually salient than the unlabeled object. Thus, models that rely directly on perceptual salience do not capture these results.

GENERAL DISCUSSION

We have presented a model that unifies cross-situational statistical approaches and intentional approaches to word learning. The model performs well in learning words from a natural corpus and also predicts a variety of behavioral phenomena reported in the word-learning literature. Previous evaluations of word-learning models have focused on either their behavioral coverage (Regier, 2005) or their performance in learning words from corpus data (Yu & Ballard, 2007), but, to our knowledge, our study is the first systematic attempt to evaluate models on both criteria.

Our model operates at the "computational theory" level of explanation (Marr, 1982). It describes explicitly the structure of a learner's assumptions in terms of relationships between observed and unobserved variables. Thus, in defining our model, we have made no claims about the nature of the mechanisms that might instantiate these relationships in the human brain. This kind of ideal-observer analysis is only one part of a full account of early word learning, and many other computational models can provide insights into different aspects of this process (Colunga & Smith, 2005; Gold & Scassellati, 2007; Li, Zhao, & MacWhinney, 2007; Regier, 2005).

The success of our model supports the hypothesis that specialized principles may not be necessary to explain many of the smart inferences that young children are able to make in learning words. Instead, in some cases, a representation of speakers' intentions may suffice. Our model is only a first step, but we hope that this work will inspire future modelers to use intentional inference to unite the rich variety of information available to young word learners.

Acknowledgments—The authors gratefully acknowledge Roberta Golinkoff, Kathy Hirsch-Pasek, Fei Xu, and Chen Yu for valuable discussions. The first author was supported by a Javits Fellowship. The second and third authors were supported by grants from the Air Force Office of Scientific Research, the Office of Naval Research, and the James S. McDonnell Foundation Causal Learning Collaborative Initiative.

REFERENCES

Baldwin, D. (1993). Early referential understanding: Infants' ability to recognize acts for what they are. *Developmental Psychology*, *29*, 832–843.

Bloom, P. (2002). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Brown, P.F., Pietra, S.D., Pietra, V.J.D., & Mercer, R.L. (1994). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, *19*, 263–311.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carlin, B.P., & Louis, T.A. (1996). *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall.

Clark, E.V. (1988). On the logic of contrast. *Journal of Child Language*, *15*, 317–335.

Clark, E.V. (2002). *First language acquisition*. Cambridge, England: Cambridge University Press.

Colunga, E., & Smith, L.B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*, 347–382.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Gold, K., & Scassellati, B. (2007, August). *A robot that uses existing vocabulary to infer non-visual word meanings from observation*. Paper presented at the annual meeting of the Cognitive Science Society, Nashville, TN.

Golinkoff, R.M., Hirsh-Pasek, K., Bailey, L.M., & Wenger, N.R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*, 99–108.

Golinkoff, R.M., Mervis, C.B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 125–155.

Levy, R. (2007). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *31*, 581–612.

Locke, J. (1964). *An essay concerning human understanding*. Cleveland, OH: Meridian Books. (Original work published 1690)

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Vol. 2: The database* (3rd ed.). Mahwah, NJ: Erlbaum.

Marinari, E., & Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters*, *19*, 451–455.

Markman, E.M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Markman, E.M., & Wachtel, G.F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Markman, E.M., Wasow, J.L., & Hansen, M.B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*, 241–275.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*, 813–815.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Henry Holt.

Mervis, C.B., & Bertrand, J. (1994). Acquisition of the Novel Name–Nameless Category (N3C) principle. *Child Development*, *65*, 1646–1662.

Plunkett, K., Sinha, C., Møller, M.F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, *4*, 293–312.

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*, 819–865.

St. Augustine. (1963). *The confessions of St. Augustine* (R. Warner, Trans.). New York: Penguin Books. (Original work published 397)

Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Smith, L. (2000). Learning how to learn words: An associative crane. In R.M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L.B. Smith, A.L. Woodward, N. Akhtar, et al. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). New York: Oxford University Press.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*, 729–742.

Xu, F. (2002). The role of language in acquiring object concepts in infancy. *Cognition*, *85*, 223–250.

Xu, F., & Tenenbaum, J.B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272.

Yu, C., & Ballard, D. (2007). A unified model of word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165.

Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.

(Received 12/3/07; Revision accepted 9/30/08)

SUPPORTING INFORMATION

Additional Supporting Information may be found in the on-line version of this article:

Technical Appendix
Word-Learning Model Code

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.